

# **Newborn Screening for Primary Congenital Hypothyroidism: Estimating Test Performance at Different TSH Thresholds**

Rachel L Knowles<sup>1</sup>, Juliet Oerton<sup>1</sup>, Timothy Cheetham<sup>2</sup>, Gary Butler<sup>3</sup>, Christine Cavanagh<sup>4</sup>, Lesley Tetlow<sup>5</sup>, Carol Dezateux<sup>6</sup>

## **AUTHOR AFFILIATIONS**

<sup>1</sup> Life Course Epidemiology and Biostatistics, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; <sup>2</sup> Newcastle University and Department of Paediatric Endocrinology, Royal Victoria Infirmary, Newcastle-upon-Tyne, United Kingdom; <sup>3</sup> Paediatric and Adolescent Endocrinology, University College London Hospitals NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health, London, United Kingdom; <sup>4</sup> National Newborn Blood Spot Screening Programme, Public Health England, London, United Kingdom; <sup>5</sup> Department of Clinical Biochemistry, Manchester University NHS Foundation Trust, Manchester, United Kingdom; <sup>6</sup> Centre for Primary Care and Public Health, Barts and the London School of Medicine and Dentistry, Queen Mary University London, United Kingdom.

## **PRECIS**

Newborn bloodspot screening for primary congenital hypothyroidism was assessed at different test thresholds. Lowering the national threshold should improve performance of the UK screening programme.

*Knowles, et al*

**CORRESPONDENCE TO :**

Dr Rachel L Knowles  
Life Course Epidemiology and Biostatistics  
Population Policy and Practice Programme  
Great Ormond Street Institute of Child Health  
University College London  
London WC1N 1EH, United Kingdom.  
Tel: +011 44 20 7905 2278  
Email: rachel.knowles@ucl.ac.uk

**DISCLOSURE SUMMARY:** The authors have nothing to disclose.

**KEYWORDS:** newborn bloodspot screening; national surveillance; pediatric

**SOURCE OF FUNDING:** Public Health England

**ACKNOWLEDGEMENTS**

We are grateful to the British Thyroid Foundation, British Paediatric Surveillance Unit, British Society of Paediatric Endocrinology and Diabetes, UK Newborn Screening Laboratory Network and the NHS Newborn Screening Blood Spot Programme for supporting this study. We would also like to thank Paul Griffiths, Cathy Coppinger, Lucy Pylypiw, Tessa Morgan and Jamili Miah for their contributions. This work uses data provided by patients and collected by the NHS as part of their care and support.

**WORD COUNT:** 3629 (main); 250 words (abstract); 33 references

## Abstract

**Context:** Active surveillance of primary congenital hypothyroidism (CH) in a multi-ethnic population with established newborn bloodspot screening.

**Objective:** To estimate performance of newborn screening for CH at different test thresholds; to calculate incidence of primary CH.

**Design:** Prospective surveillance undertaken from June 2011 to June 2012 with three-year follow-up of outcomes. Relative likelihood ratios (rLRs) estimated to compare bloodspot thyroid-stimulating hormone (TSH) test thresholds of 6mU/L and 8mU/L, with the nationally recommended standard of 10mU/L for a presumptive positive result.

**Setting:** UK National Health Service

**Patients:** Clinician notification of children aged under five years investigated following clinical presentation or presumptive positive screening result.

**Main outcome measure(s):** Permanent primary CH status determined by clinician report of continuing thyroxine requirement at three-year follow-up.

**Results:** 629 newborns (58.3% girls; 58.7% white ethnicity) were investigated following presumptive positive screening result and 21 children (52.4% girls; 52.4% white) after clinical presentation; 432 remained on treatment at three-year follow-up. Permanent CH incidence was 5.3 (95%CI 4.8, 5.8) per 10,000 infants. Using locally-applied thresholds, sensitivity, specificity and positive predictive value were 96.76%, 99.97% and 66.88% respectively. Compared with TSH threshold of 10mU/L, positive rLRs for 8mU/L and 6mU/L were 1.20 (95%CI 0.82, 1.75) and 0.52 (95%CI 0.38, 0.72), and negative rLRs 0.11 (95%CI 0.03, 0.36) and 0.11 (95%CI 0.06, 0.20) respectively.

**Conclusions:** Screening programme performance is good, however a TSH threshold of 8mU/L appears superior to the current national standard (10mU/L) and requires further evaluation. Further research should explore the implications of transient CH for screening policy.

# Introduction

Primary congenital hypothyroidism (CH) affects around 1 in 2000 children born in the United Kingdom (UK) each year. It is estimated that 8-28% of children presenting clinically will develop severe intellectual disability, defined as an IQ <70.<sup>1</sup> Newborn screening to identify those with CH enables timely thyroxine replacement therapy and potentially prevents or mitigates this disability.<sup>2</sup> Newborn screening was introduced in the UK in 1981 and is currently based on whole blood thyroid stimulating hormone (TSH) concentrations measured in dried bloodspots collected 5 days postnatally.<sup>3</sup> Secular increases in the proportion of babies with presumptive positive screening results<sup>1</sup> may reflect a number of factors, including increasing ethnic diversity<sup>4</sup>, changes in maternal iodine status<sup>5,6</sup>, and reduction over time in the lower limit of TSH threshold used to define a presumptive positive result reflecting technological advances in laboratory measurement.<sup>1,7</sup>

The UK national standards recommend confirmatory diagnostic testing in all infants with screening bloodspot TSH (whole blood)  $\geq 20\text{mU/L}$ , or  $\geq 10\text{mU/L}$  after repeat testing for borderline results (Table 1), however in practice there is variability in TSH thresholds for technological and historical reasons<sup>1</sup> and the current performance of the national programme has not been appraised. At the time of this study, 12 of the 16 UK newborn screening laboratories employed a TSH threshold below the recommended national standard, largely due to concerns about false negative results. This provided a rare opportunity to evaluate screening test performance at different TSH thresholds within an existing national programme involving a multi-ethnic population of 700,000 births per year.

We carried out a prospective UK-wide active surveillance study to identify all confirmed diagnoses of primary CH in children aged under five years, regardless of screening results. As CH may be transient in the early years, we obtained reports of outcomes in notified infants after diagnosis and with expert advice from paediatric endocrinologists, developed and applied standardised criteria for defining 'confirmed' and 'probable' CH status from

clinician reports at three year follow-up. To inform future screening policy, we assessed incremental changes in the detection rate, false positive rate and likelihood ratio of two alternatives to the current recommended threshold of  $\geq 10\text{mU/L}$  for defining a presumptive positive screening result ( $\geq 8\text{mU/L}$  or  $\geq 6\text{mU/L}$ ).

## Materials and Methods

### Ascertainment of cases

Children with newly diagnosed CH were identified by active surveillance through the British Paediatric Surveillance Unit (BPSU) national clinical paediatric surveillance system and through a concurrent laboratory reporting system, involving all 16 UK newborn screening laboratories. Secondary and tertiary care paediatricians, and laboratory directors, notified monthly all children meeting the reporting case definition (Table 1) and full clinical details were obtained from the notifying clinician and laboratory using online questionnaires.<sup>8</sup> Laboratory and clinician notifications were matched using birth date, National Health Service (NHS) number or equivalent, sex and postcode district. If a case had not been reported by both sources, we asked paediatricians or laboratories to complete a questionnaire or provide further clinical details.

### Follow up and outcome adjudication

All children were followed up annually using online or postal questionnaires sent to clinicians until one of the following endpoints: completion of three years follow up, death, CH confirmed, discharged from/lost to clinician care. Collected data included details of screening and diagnostic test results, clinical presentation and management.

An independent expert panel, comprising two paediatric endocrinologists and one screening laboratory director, reviewed every child's de-identified data to determine (1) eligibility for study inclusion and (2) outcome at three years. Children had *confirmed permanent* CH at

three years if a persisting requirement for thyroxine was confirmed by a trial off therapy (withdrawal of thyroxine replacement therapy and re-evaluation of thyroid function tests to confirm or exclude CH), OR radioisotope or ultrasound scan results confirming thyroid agenesis or ectopic thyroid, OR continuing requirement for 'high dose' levothyroxine ( $\geq 50\text{mcg}$  per day) indicated by regular review of thyroid function by paediatricians. Children had *probably permanent* CH if confirmation of CH as defined above was absent but the clinician was continuing levothyroxine at final follow-up. Children were confirmed *not CH* if not on treatment by three-year follow-up; children who had a period on levothyroxine before treatment was discontinued, following a trial off therapy or other clinical evaluation (not specified by the clinician), were defined as *transient CH*.

## Test performance, incidence and standardised population

Incidence of permanent CH diagnosis in UK infants was estimated, using monthly live birth data from the Office for National Statistics (ONS)<sup>9</sup>, National Registrations Scotland<sup>10</sup> and Northern Ireland Registration and Statistics Authority, for England and Wales, Scotland and Northern Ireland respectively.<sup>11</sup> Ethnic groupings were White, Asian, Black, Mixed and Other (UK Census 2011 categories).<sup>12</sup> Incidence rates (IR) by sex, gestation and ethnicity were estimated *for England only* using ONS live birth data ( $n=693,748$  live births<sup>13</sup>); incidence rate ratios (IRR) were estimated for comparison with reference categories (Table 2). Analyses comparing screen thresholds used standardised English live birth data<sup>13</sup>, adjusted for between-laboratory population differences by sex, gestation and ethnicity that could influence screening outcomes.<sup>4</sup>

Performance of the UK newborn screening programme, in detecting confirmed/probably permanent CH, was evaluated from 2011 to 2012 ( $n=813,087$  live births), after excluding four infants diagnosed before screening and four with indeterminate outcome. In separate sensitivity analyses, (i) children with probably permanent CH were assigned to the 'not CH'

category and (ii) infants with indeterminate outcome were assumed to be all true or all false positive cases.

Laboratories reported actual TSH values for positive screen results, and all values below the local threshold as 'screen negative', therefore a continuous receiver operating curve (ROC) could not be plotted to compare thresholds. Instead three groups of English screening laboratories were defined by the lower TSH threshold each used, Group 1 (n=5 laboratories; TSH $\geq$ 5 or  $\geq$ 6mU/L), Group 2 (n=3; TSH $\geq$ 8mU/L) and Group 3 (n=4; TSH $\geq$ 10mU/L), and screening performance compared between groups. As populations served by the laboratories differed in ethnic preterm birth rate profiles, we directly standardised populations<sup>14</sup> for comparison. We applied screen positive rates by sex, ethnicity and gestation from each laboratory group to the English population of 693,748 live births and adjusted the results to a population of 100,000 infants. The trade-off between sensitivity and specificity for each laboratory group was plotted on a ROC of sensitivity versus false positive rate (1-specificity).

Test performance at different TSH thresholds was compared by estimating positive (rLR+) and negative (rLR-) relative likelihood ratios, using the method described by Hayen<sup>15</sup> and assuming that a threshold of TSH $\geq$ 6mU/L and TSH $\geq$ 8mU/L were replacement screening thresholds for TSH $\geq$ 10mU/L. Where the rLR+ for the new threshold is  $>1$  compared with the current threshold, this indicates that the new threshold is *more likely to correctly* assign a positive screen result to a child *with* CH, while a rLR- for the new threshold of  $<1$  indicates the new threshold is *less likely to incorrectly* assign a positive screen result to a child *without* CH.

## Statistical Analysis

Statistical analyses were performed using StataSE13 (StataCorp, TX). Research ethics approval (Cambridge South REC; 11/EE/0152) and Section 251 support were obtained (ECC 3-04(k)/2011).

## Results

There were 518 notifications from clinicians and 704 from laboratories. We excluded 75 duplicates (cases reported twice through the same source), 118 diagnosed before 01 July 2011 or not meeting our case definition (including non-UK births), and 41 cases that remained unverified as sufficient clinical details were not provided. Of those remaining, 338 were 'matched' notifications reported by both laboratory and clinician sources.

All further analyses are based on 650 individual cases reported to the study during 12 months between 01 July 2011 and 30 June 2012, comprising 629 children investigated after a presumptive positive newborn screen result and 21 reported as 'clinically detected' by paediatricians. The total population screened was 813087 (Table 3), of which 1.5% babies were born <32 weeks gestation and followed the 'preterm' screening pathway which included a repeat whole blood sample (Table 1).

Children reported following a presumptive positive screen result were more likely to be girls (n=367; 58.3%) and of White (n=369; 58.6%) or Asian (n=128; 20.3%) ethnicity. Fifty (7.9%) babies were born <32 weeks gestation. Twelve deaths occurred and all were associated with prematurity or comorbidities; one infant was being treated for CH, 10 did not have CH, and one died before diagnostic tests were completed.

Of 21 clinically detected children, 11 were girls, 11 were of White ethnicity and six were born <32 weeks gestation; one death occurred which was unrelated to CH. CH was not suspected at newborn screening in 17 ('screen negative') of these children and four were referred for investigation *before* the screening results were available; we refer to all of these as 'clinically detected' cases as they were not identified through the newborn screening pathway.



## Diagnostic outcomes

### *Infants with a presumptive positive screen*

At initial clinical referral (ICR), 488 (77.6%) of 629 children were diagnosed with CH and commenced levothyroxine; CH was excluded in 137 (21.8%) infants (Figure 1a). Diagnostic tests remained incomplete in four children (indeterminate outcome), one of whom died.

By three years of age, 295 children had confirmed permanent CH, of whom 33 had a trial off therapy, 165 had scan confirmation of agenesis or ectopic thyroid, and 97 required high dose levothyroxine. A further 123 children had probably permanent CH. CH was excluded in 207 children (trial off therapy (n=58) or other clinical evaluation [n=149]; Figure 1a), of whom 70 received thyroxine for <3 years (transient CH). Of 50 screen positive babies born <32 weeks gestation, 16 had confirmed/probably permanent CH at 3 years.

### *Clinically detected children*

At ICR, 20 of 21 children were diagnosed with CH and started levothyroxine; CH was excluded before treatment in one (Supplementary Table S4). Six children were born <32 weeks gestation and had a repeat screen, and five were born between 32 weeks and <37 weeks gestation. Four children suspected *before* screening had comorbidities and/or family history and all remained on treatment at 3 years. Two of these babies had a blood spot TSH  $\geq 10$  mU/L screening (20 and 40 mU/L) however they were referred before these screening results were reported.

By three year follow-up, four children had confirmed permanent CH; three of these had a trial off therapy and one required high dose thyroxine (Figure 1b). These children presented with a congenital anomaly, family history or prolonged jaundice; all had bloodspot TSH <8 mU/L and started levothyroxine by age three months. Ten children had probably permanent CH at 3 years; all bloodspot TSH were  $\leq 8$  mU/L (and <6 mU/L in seven children). Four of six babies born <32 weeks gestation had confirmed/probably permanent CH at 3

years. CH was excluded in seven children by 3 years; six had transient CH (confirmed by trial off therapy [n=3] or other clinical evaluation [n=3]) and one never started treatment.

## Incidence of CH

There were 432 infants born between 1<sup>st</sup> July 2011 and 30<sup>th</sup> June 2012 and subsequently diagnosed with confirmed/probably permanent CH (418 with a positive screen [Figure 1a] and 14 clinically detected [Figure 1b]); no child presenting after age one year had confirmed/probably permanent CH. UK birth prevalence was 5.3 (95% CI: 4.8, 5.8) per 10,000 live births.

Incidence of permanent CH by sex, gestation and ethnicity was estimated for English live births (Table 2). Incidence of permanent CH was significantly higher for girls (IRR 1.5 [1.2, 1.8]), and for infants born before 32 weeks gestation compared with those born at or after 32 weeks (IRR 3.7 [2.2, 5.9]). Compared to children of White ethnicity (IR 4.5 [95%CI: 4.0, 5.1]), children of Asian and Chinese ethnicity had a significantly higher incidence of permanent CH (IRR: Asian 2.5 [1.9, 3.2], Chinese 4.2 [1.7, 8.7] respectively), while children of Black ethnicity had lower incidence (IRR 0.4 [0.1, 0.8]).

## Screening programme performance

Evaluation of UK-wide screening programme performance, using locally-determined TSH thresholds, demonstrated high sensitivity 96.76% (95% CI: 94.62%, 98.22%) and specificity 99.97% (95% CI: 99.97%, 99.98%), for a PPV of 66.88% (95% CI: 63.04%, 70.56%; Table 3). The likelihood ratio for a positive screen result (LR+), or the odds of a child having permanent CH if the screening test is positive, was high at 3799. Sensitivity analyses assigning children with probable CH to 'not CH' demonstrated similar screening sensitivity (95.47% [95%CI 92.54%, 97.28%]) and specificity (99.96% [95%CI 99.95%, 99.96%]) but lower PPV (47.20% [95%CI 43.32%, 51.12%]). Sensitivity analyses re-assigning infants with indeterminate outcomes did not significantly alter test performance (data not shown).

## Screening performance at different bloodspot TSH thresholds

Screening performance at three TSH thresholds used by different groups of English laboratories ( $\geq 6$  mU/L,  $\geq 8$  mU/L,  $\geq 10$  mU/L) was compared for a population of 100,000 English live births standardised by sex, gestation and ethnicity (Supplementary Table S1). At TSH thresholds lower than the national standard ( $\geq 10$  mU/L), the sensitivity and false positive rate increased, and PPV decreased, being 62.2% at  $\geq 6$  mU/L.

A plot of sensitivity and specificity for each laboratory group (Figure 2) suggests that the optimal TSH threshold lies between  $\geq 6$  and  $\geq 10$  mU/L. This was supported by the positive (rLR+) and negative (rLR-) relative likelihood ratios estimated for screening test performance at TSH  $\geq 6$  mU/L and TSH  $\geq 8$  mU/L, compared with TSH  $\geq 10$  mU/L (Table 4). Compared with a TSH threshold  $\geq 10$  mU/L, the rLR+ value of  $\geq 8$  mU/L was  $>1$  and rLR- was  $<1$ . As 95% CI for rLR+ included 1, we cannot exclude the possibility that TSH  $\geq 8$  mU/L does not differ significantly from the current national standard (TSH  $\geq 10$  mU/L)<sup>15</sup>, nevertheless these results suggest that the NPV for  $\geq 8$  mU/L is superior to  $\geq 10$  mU/L without appreciable reduction in PPV. In contrast, the rLR+ and rLR- were  $<1$  for TSH  $\geq 6$  mU/L, suggesting the PPV at TSH  $\geq 6$  mU/L is inferior to  $\geq 10$  mU/L. Sensitivity analyses reassigning 'probably permanent' CH cases to 'not CH' did not change the rLR- values but, compared with TSH  $\geq 10$  mU/L, the rLR+ for TSH  $\geq 8$  mU/L reduced to 1.02 (95% CI 0.76, 1.37) and the TSH  $\geq 6$  mU/L remained  $<1$  (rLR+ 0.66 [95% CI 0.51, 0.84]).

Had all English laboratories in this study been using TSH  $\geq 10$  mU/L, 10 children with confirmed permanent CH screened in laboratories using thresholds between TSH  $\geq 6$  mU/L and  $<10$  mU/L, might have been 'missed' (Supplementary Table S3). At a threshold TSH  $\geq 8$  mU/L, six children would have been screen positive, while the remaining four infants had bloodspot TSH values  $<8$  mU/L (one had a congenital syndrome associated with CH, and three had dysghormonogenesis with normal scans).

## Discussion

In a prospective UK-wide study of CH using active reporting by clinicians and newborn screening laboratories, we found that only two-thirds of those with an initial diagnosis of CH following a presumptive positive screening result continued to require thyroxine treatment three years later. We estimate that, in England, CH incidence is higher in girls, babies born <32 weeks gestation or of Asian or Chinese ethnicity, and overall higher than before screening was introduced. Our evaluation of screening programme performance demonstrated that the UK programme has high sensitivity, specificity and positive predictive value. Importantly we have shown that replacing the national recommended threshold of  $TSH \geq 10 \text{ mU/L}$  with a lower threshold of  $TSH \geq 8 \text{ mU/L}$  would likely result in improved test performance and identify infants who are currently detected at thresholds below the current recommended threshold, without concomitant increase in false positive screening results. We found no substantial advantage in test performance using a threshold of  $TSH \geq 6 \text{ mU/L}$ . Importantly, these thresholds are in relation to the UK screening programme, in which the newborn bloodspot is taken at 5 days of age, and therefore these thresholds may not apply to programmes that perform screening earlier or later.

We identified a contemporary incidence of CH that is approximately double that reported in the UK population before newborn screening was introduced,<sup>16</sup> similar to the increase noted with introduction of screening in other European and north American countries.<sup>1,17</sup> This rise may be related to changes in population demographics<sup>1,18</sup> and, in the UK, ethnic variation in thyroid physiology has been proposed as underlying the growth in screen-detected cases.<sup>4</sup> Schoen has highlighted variations by sex and ethnicity in the population distribution of mild and severe CH, which may reflect different causes.<sup>19</sup> As maternal iodine insufficiency leads to raised newborn TSH, higher rates of positive screen results may be partly due to increased prevalence of insufficiency amongst UK women<sup>5,6</sup>; this merits further investigation.

Lower TSH thresholds may also contribute to the observed increase in CH incidence through increased detection of 'transient', mild or subclinical CH; the implications for neurodevelopmental outcomes and need for lifelong treatment are less clear for these children.<sup>20,21</sup> Alm reported that children with subclinical CH, defined as raised TSH without other symptoms and signs of CH, had similar neurodevelopmental outcomes to unaffected controls.<sup>22</sup> More recently, Lain<sup>23</sup> showed that children with marginally raised newborn TSH results, below the levels indicated for treatment in the Australian programme, perform less well educationally than children with treated CH or with negative screen results at lower TSH levels, suggesting the potential for subtle cognitive impairment due to mild CH.

However lower thresholds lead to significant increases in false positive rates.<sup>20,24</sup> Korada<sup>24</sup> reported 126% increase in false positive rate on lowering the TSH threshold from 20mU/L to 6mU/L. Furthermore the investigation of false positive results increases the costs of screening<sup>25,26</sup>, and can lead to persisting anxiety in parents even after exclusion of CH.<sup>27,28</sup> Children treated for mild or severe CH may experience reduced quality of life<sup>29</sup> compared with unaffected peers, and neurodevelopment may be adversely affected by frequent monitoring which raises concerns for parents and children.<sup>30</sup> The harms of over-investigation and over-treatment, including the continuing treatment lifelong in children for whom CH is not confirmed, are significant and should not be ignored when evaluating newborn screening.

As in our study, Ford found that US newborns who were identified as presumptive positive CH on the first bloodspot were more likely to be girls and to have permanent CH than those who were referred on a repeat test.<sup>17</sup> He suggests that the first test may identify infants with prenatal onset of CH due to agenesis or ectopic thyroid, which are more common in girls.

False negative rates in our study were higher than those reported previously<sup>7</sup>, however these are likely to have been underestimated in previous studies which used less reliable methods for ascertaining clinically presenting cases. Two studies<sup>31,32</sup> using multiple sources to capture false negative cases reported rates of 0.1 and 0.3 per 100,000 respectively, which

compares with our rate of 1.1 per 100,000 infants screened. These false negative or 'missed' cases underline the importance of checking thyroid function in older infants who present with clinical manifestations that may indicate hypothyroidism, as inevitably not all cases can be detected by population screening programmes even when very low TSH thresholds are used.

Important strengths of our observational study were the complete national coverage of a large population of over 800,000 newborns in which all screening laboratories were using AutoDELFIA (Perkin Elmer) technology, and high ascertainment and follow up rates. Moreover the high rate of ascertainment of clinically presenting screen-negative cases, permits reliable estimates of screening programme performance. Nevertheless the laboratory source was essential for achieving complete ascertainment, as some paediatricians did not report all cases identified as presumptive positive by screening. Paediatricians were more likely to report a screen positive infant if they started treatment, whereas laboratory staff reported all screen positive infants regardless of subsequent treatment decisions.

Although differences in screening thresholds between laboratories introduced variability into our estimates of sensitivity and specificity for the screening programme as a whole, we were able to take advantage of these to evaluate the influence of bloodspot TSH thresholds on screening performance. Furthermore using direct population standardization, we ensured that differences between the three laboratory groups, including population ethnicity, were accounted for in our comparative analyses.

Unlike many previous studies, we undertook follow-up to three years after initial referral and obtained information about re-evaluation and confirmatory tests throughout this period to inform the final assignment of diagnostic outcome. Had we relied upon the diagnosis at onset of therapy, the estimated number of CH cases would have been 16% higher. However as this was an observational study, clinicians completed questionnaires using only the data that was routinely available in medical records therefore information about the reasons for

clinical decisions was limited. We assumed at three-year follow-up that children who continued on a levothyroxine dose of  $<50\text{mcg}$  per day without re-evaluation or scan confirmation had probably permanent, rather than transient, CH: should this assumption prove incorrect, this would result in under-estimation of transient cases and over-estimation of probably permanent CH cases.

Our study demonstrates that, in the UK, 30% of children with a presumptive positive screen continue long-term on thyroxine treatment without a trial off therapy or other confirmation of permanent CH. This underlines the need for a more active approach to re-evaluating CH diagnosis in all children around 2-3 years of age to avoid lifelong levothyroxine in children who do not require it.

Analysis of the trade-off between sensitivity and specificity at screening thresholds of  $\geq 6$ ,  $\geq 8$  and  $\geq 10\text{mU/L}$  suggests that the optimal TSH threshold is likely to be around  $8\text{mU/L}$  for infants screened at 5 days of life. A reduction in screen test thresholds that completely avoids 'missed' cases is not feasible and would likely result in a higher numbers of children undergoing unnecessary investigation and treatment for CH. Most children in our study who presented clinically after a negative screen result were identified through investigation of prolonged jaundice or comorbidities.

Existing cost-benefit analyses for the UK screening programme for CH are based on preventing severe intellectual disability<sup>33</sup>, however there is no clear evidence that these benefits apply to all types of CH, including children identified at lower screen thresholds. Further investigation of the natural history and benefits of treating mild, transient or subclinical CH is essential to confirm the benefit or otherwise of extending the current screening programme to detect such cases. Further research is essential to understand the characteristics and outcomes for infants with mild or transient CH in order to offer an effective population screening programme that appropriately balances the benefit of early diagnosis against the harms of over-investigation and over-treatment.

## References

1. Pollitt RJ. Evidence or enthusiasm? Why yields from UK newborn screening programmes for congenital hypothyroidism are increasing. *Arch Dis Child* 2016;101:120-3.
2. Léger J, Olivieri A, Donaldson MD, et al. European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. *J Clin Endocrinol Metab* 2014;99:363-84.
3. Newborn blood spot screening: programme overview (2013). (Accessed 25 July 2017 at: [www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview](http://www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview))
4. Peters C, Brooke I, Heales S, et al. Defining the Newborn Blood Spot Screening Reference Interval for TSH: Impact of Ethnicity. *J Clin Endocrinol Metab* 2016;101:3445-9.
5. Vanderpump MP, Lazarus JH, Smyth PP, et al. Iodine status of UK schoolgirls: a cross-sectional survey. *Lancet* 2011;377:2007-12.
6. Rayman M, Sleeth M, Walter A, Taylor A. Iodine deficiency in UK women of childbearing age. *Proc Nut Soc* 2008;67:E399.
7. Knowles RL, Olafsdottir F. Initial Clinical Referral Standards after Newborn Screening for Congenital Hypothyroidism: Final Report of the UK Newborn Screening Programme Centre Expert Working Group and Systematic Evidence Review 2010-2011. *J Open Public Health Data* 2013;1(1):e5:e5.
8. REDCap software, Vanderbilt University. URL: [www.redcap.org](http://www.redcap.org)
9. Characteristics of Birth 2, England and Wales. In: ONS, ed.2012.
10. Scotland) NNR. Monthly Data on Births and Deaths Registered in Scotland. Scotland: National Records Scotland; 2017 16th February 2017.
11. Resident Live Births in Northern Ireland by Month of Registration, 2006-2016. 2016. (Accessed 07 March 2017 at: [www.nisra.gov.uk/archive/demography/vital/births/MonthlyBirths.xls](http://www.nisra.gov.uk/archive/demography/vital/births/MonthlyBirths.xls))
12. Ethnic Group. Statistical bulletin - 2011 Census: Key Statistics for England and Wales, March 2011. United Kingdom: Office for National Statistics; 2011:10-2.
13. Macfarlane A, Dodwell M, Gibson R, Newburn M, Dattani N, Cortina-Borja M. Births and their outcome: analysing the daily, weekly and yearly cycles and their implications for the NHS. NIHR Research Project reference HS&DR - 12/136/93. (Accessed 17 June 2018 at: <https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/1213693/#/>)
14. Eayres D. *APHO Technical Briefing 3: Commonly used public health statistics and their confidence intervals*. London, UK: Public Health England; 2008.
15. Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epidemiol* 2010;63:883-91.
16. Hulse JA. Outcome for congenital hypothyroidism. *Arch Dis Child* 1984;59:23-9.
17. Ford G, LaFranchi SH. Screening for congenital hypothyroidism: a worldwide view of strategies. *Best Pract Res Clin Endocrinol Metab* 2014;28:175-87.
18. Harris KB, Pass KA. Increase in congenital hypothyroidism in New York State and in the United States. *Mol Genet Metab* 2007;91:268-177.
19. Schoen EJ, Clapp W, To TT, Fireman BH. The key role of newborn thyroid scintigraphy with isotopic iodide (123I) in defining and managing congenital hypothyroidism. *Pediatrics* 2004;114:e683-8.
20. Langham S, Hindmarsh P, Krywawych S, Peters C. Screening for congenital hypothyroidism: comparison of borderline screening cut-off points and the effect on the number of children treated with levothyroxine. *Eur Thyroid J* 2013;2:180-6.



21. Corbetta C, Weber G, Cortinovis F, et al. A 7-year experience with low blood TSH cutoff levels for neonatal screening reveals an unsuspected frequency of congenital hypothyroidism (CH). *Clin Endocrinol* 2009;71:739-45.
22. Alm J, Hagenfeldt L, Larsson A, Lundberg K. Incidence of congenital hypothyroidism: retrospective study of neonatal laboratory screening versus clinical symptoms as indicators leading to diagnosis. *BMJ* 1984;289:1171-5.
23. Lain SJ, Bentley JP, Wiley V, et al. Association between borderline neonatal thyroid-stimulating hormone concentrations and educational and developmental outcomes: a population-based record-linkage study. *Lancet Diabetes Endocrinol* 2016;4:756-65.
24. Korada SM, Pearce M, Ward Platt MP, et al. Difficulties in selecting an appropriate neonatal thyroid stimulating hormone (TSH) screening threshold. *Arch Dis Child* 2010;95:169-73.
25. Pollitt. Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assess* 1997;1:i-iv, 1.
26. Heather NL, Hofman PL, de Hora M, Carll J, Derraik JG, Webster D. Evaluation of the revised New Zealand national newborn screening protocol for congenital hypothyroidism. *Clin Endocrinol* 2016.
27. Fyro K, Bodegaard G. Difficulties in Psychological Adjustment to a New Neonatal Screening Programme. *Acta Paediatrica* 1988;77:226-31.
28. Hewlett J, Waisbren SE. A review of the psychosocial effects of false-positive results on parents and current communication practices in newborn screening. *J Inherit Metab Dis* 2006;29:677-82.
29. van der Sluijs Veer L, Kempers MJE, Last BF, Vulsma T, Grootenhuys MA. Quality of Life, Developmental Milestones, and Self-Esteem of Young Adults with Congenital Hypothyroidism Diagnosed by Neonatal Screening. *J Clin Endocrinol Metab* 2008;93:2654-61.
30. Kohler B, Schnabel D, Biebermann H, Gruters A. Transient congenital hypothyroidism and hyperthyrotropinemia: normal thyroid function and physical development at the ages of 6-14 years. *J Clin Endocrinol Metab* 1996;81:1563-7.
31. Pharoah PO, Madden MP. Audit of screening for congenital hypothyroidism. *Arch Dis Child* 1992;67:1073-6.
32. Ray M, Muir TM, Murray GD, Kennedy R, Girdwood RW, Donaldson MD. Audit of screening programme for congenital hypothyroidism in Scotland 1979-93. *Arch Dis Child* 1997;76:411-5.
33. Grosse SD, Van Vliet G. Prevention of intellectual disability through screening for congenital hypothyroidism: how much and at what level? *Arch Dis Child* 2011;96:374-9.

**Table 1: Screening and Surveillance Definitions\***

**National guidelines for newborn blood spot screening:**

The first newborn bloodspot sample is taken at **5 days of age** in all babies.

Babies born at **less than 32 weeks gestation** also have a second (repeat) bloodspot sample at 28 days of age or on the day of discharge home, whichever is sooner, as immaturity may mask CH.

A **presumptive positive screening** result requiring referral for diagnostic investigation is defined as a TSH concentration of  $>20\text{mU/L}$  on the newborn blood spot (whole blood) sample; a concentration between 10 and 20 mU/L is a 'borderline' result requiring a repeat screen and diagnostic referral if the TSH level remains  $\geq 10\text{mU/L}$  in the second blood spot sample.

**Clinical referral guidelines** recommend thyroid function tests (serum TSH and free T4) to confirm the diagnosis after a presumptive positive screen as well as ultrasound and/or radio-isotope scans to determine the underlying thyroid gland abnormality.<sup>5</sup>

**Treatment** – oral thyroxine which should be initiated by 21 days of age.<sup>2,5</sup>

-----

**Reporting case definition for the surveillance study:**

Any child up to and including five years of age who, during the past month, has been referred:

- EITHER for diagnostic confirmation following a newborn screening test result suggestive of primary congenital hypothyroidism (CH),
- OR has been confirmed with a diagnosis of primary CH (known or considered likely to be present from birth), based on a serum  $\text{TSH} \geq 10\text{mU/L}$ .

**Table 2: Annual incidence of diagnosis of CH per 10,000 live births in England**

	Confirmed/ Probable CH (n)	Births (n) in England*	Incidence (95% CI) per 10,000 live births	Rate ratio <sup>§</sup> (95% CI)
<b>Sex</b>				
Male	148	338,081	4.4 (3.7, 5.1)	reference
Female	227	355,667	6.4 (5.6, 7.3)	1.5 (1.2, 1.8)
Not known	0	2592	-	-
<b>Gestation at birth</b>				
≥32 weeks	354	683,829	5.2 (4.7, 5.7)	reference
<32 weeks	19	9,919	19.2 (11.5, 29.9)	3.7 (2.2, 5.9)
Not known	2	2592	-	-
<b>Ethnicity</b>				
White	231	510,586	4.5 (4.0, 5.1)	reference
Asian	82	73,466	11.2 (8.9, 13.9)	2.5 (1.9, 3.2)
Black	6	36,264	1.7 (0.6, 3.6)	0.4 (0.1, 0.8)
Mixed	23	34,969	6.6 (4.2, 9.9)	1.5 (0.9, 2.2)
Chinese	7	3,724	18.8 (7.6, 38.7)	4.2 (1.7, 8.7)
Other	12	13,484	8.9 (4.6, 15.5)	2.0 (1.0, 3.5)
Not known	14	23,847	-	-

**Notes** \* Denominators are from 693,748 live births in England by sex, ethnicity and gestation between July 2011 and June 2012 (data provided by Professor M Cortina-Borja); the numerator is 375 probable/confirmed CH cases in England only (as these denominators were not available for Scotland, Northern Ireland and Wales). <sup>§</sup> The incidence rate ratio is estimated for the incidence rate within each category compared with the reference. **Abbreviations** CH congenital hypothyroidism; n number; CI confidence intervals

**Table 3: Performance of the UK newborn screening programme for CH, 2011-2012**

	Confirmed/Probable CH (n)	CH excluded (n)	Total
<b>Screen positive*</b>	418	207	625
<b>Screen negative*</b>	14	812448	812462
<b>Total (n)</b>	432	812655	813087

**Screening performance**

Sensitivity (%; 95%CI)	96.76%	(94.62%, 98.22%)
Specificity (%; 95%CI)	99.97%	(99.97%, 99.98%)
Positive predictive value (PPV; %; 95%CI)	66.88%	(63.04%, 70.56%)
False positive rate (%; 95%CI)	0.03%	(0.02%, 0.03%)
Likelihood ratio positive (LR+)	3799	
Likelihood ratio negative (LR-)	0.03	

\*Screen result as defined by local laboratory TSH thresholds; outcome as defined at 3 year follow-up.

**Abbreviations** CH congenital hypothyroidism; n number; CI confidence intervals

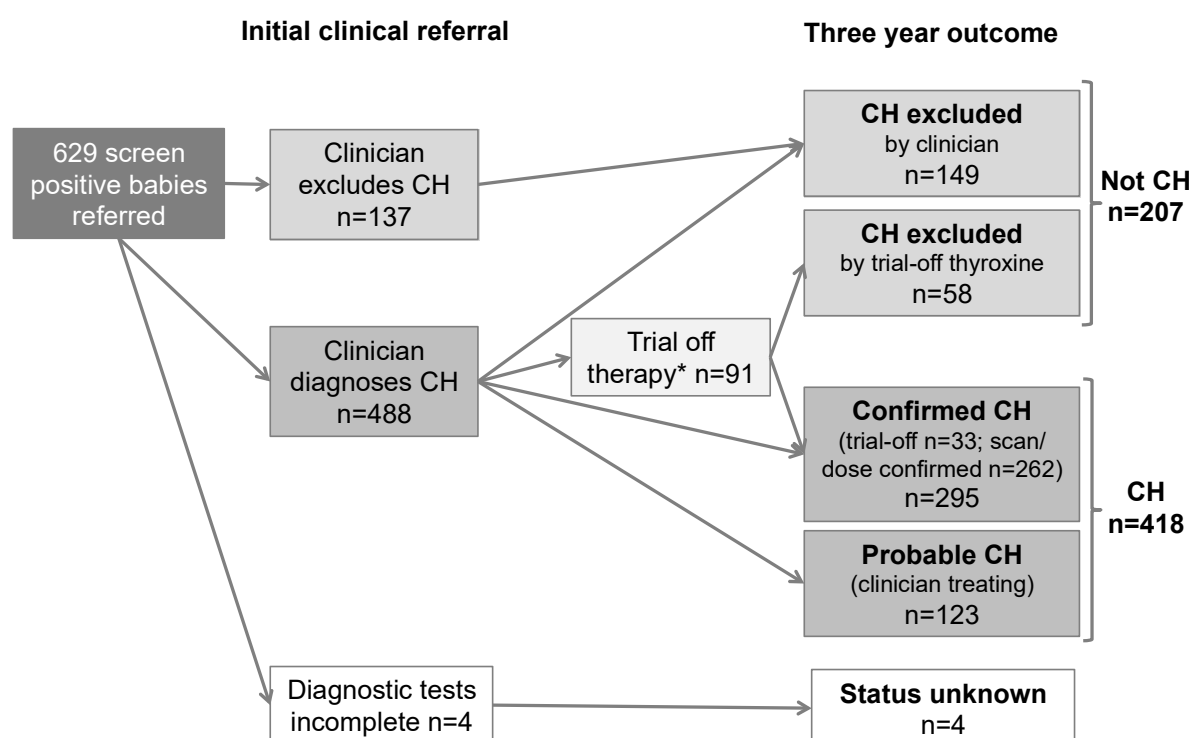
**Table 4: Relative likelihood ratios for screen thresholds replacing TSH≥10mU/L**

TSH≥6mU/L as a replacement for TSH≥10mU/L*		
	LR+	LR-
Existing test (TSH ≥10mU/L) (Groups 2&3; n=377,914)	5303.00	0.16
Replacement test (TSH ≥6mU/L) (Group 1; n=315,944)	2773.00	0.02
rLR+	0.52 (95%CI 0.38, 0.72)	
rLR-	0.11 (95%CI 0.03, 0.36)	
TSH≥8mU/L as a replacement for TSH≥10mU/L**		
	LR+	LR-
Existing test (TSH ≥10mU/L) (Group 3; n=252,028)	5632.00	0.16
Replacement test (TSH ≥8mU/L) (Groups 1&2; n=441,830)	4691.00	0.02
rLR+	1.20 (95%CI 0.82, 1.75)	
rLR-	0.11 (95%CI 0.06, 0.20)	

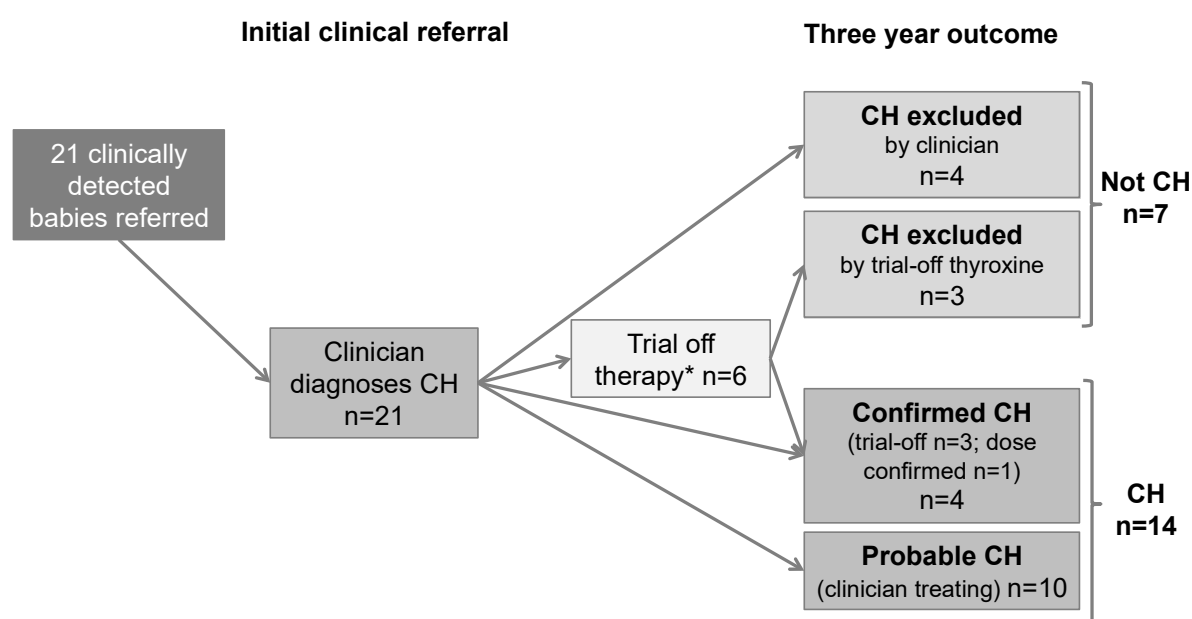
**Notes** \* Screen performance was estimated for all children in Laboratory Group 1 (TSH≥6mU/L) and compared with all children in Laboratory Groups 2 and 3 combined (using TSH≥10mU/L as the screen thresholds); \*\* Screen performance was estimated for all children in Laboratory Groups 1 and 2 combined (using TSH≥8mU/L as the screen thresholds and treating all values below this as negative) and compared with all children in Laboratory Group 3 (TSH≥10mU/L); **Abbreviations** TSH thyroid stimulating hormone; CH congenital hypothyroidism; CI confidence interval; LR+ likelihood ratio for a positive screen result; LR- likelihood ratio for a negative screen result; rLR+ positive relative likelihood ratio; rLR- negative relative likelihood ratio;

**Figure 1: Flow diagram of outcomes at initial clinical referral and three year follow-up**

**(a) for 629 babies referred as screen positive**



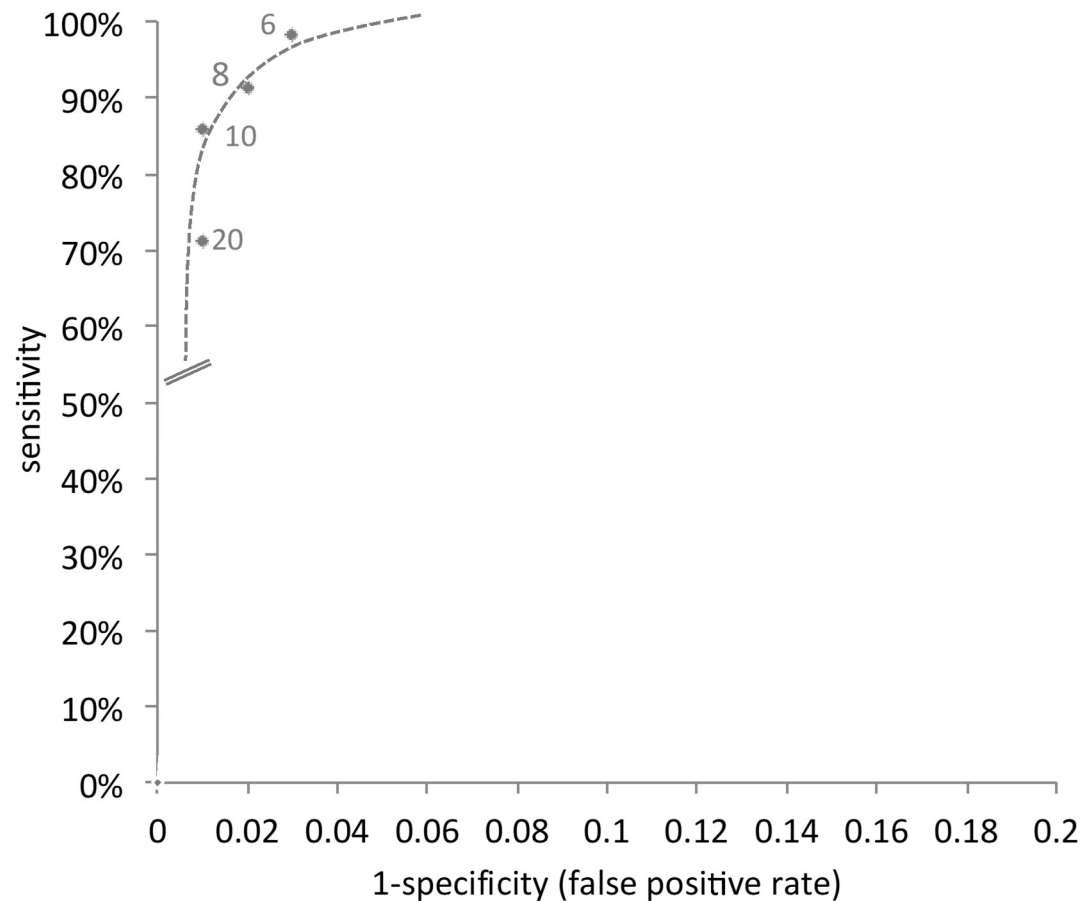
**(b) for 21 babies referred as clinically detected**



**Notes** \* Trial off therapy: a period of withdrawal of thyroxine replacement therapy to allow re-evaluation of thyroid function tests off therapy and confirmation of permanent or transient CH.

**Abbreviations** CH congenital hypothyroidism;

**Figure 2: Receiver Operating Curve by English laboratories grouped according to TSH screening thresholds used**



**Notes** The sensitivity and 1-specificity is plotted for each group of English laboratories using different screening thresholds: **6** = laboratories in **Group 1** ( $\text{TSH} \geq 6 \text{ mU/L}$ ); **8** = laboratories in **Group 2** ( $\text{TSH} \geq 8 \text{ mU/L}$ ); **10** = **Group 3** laboratories ( $\text{TSH} \geq 10 \text{ mU/L}$ ); **20** = combined performance of all laboratories at this threshold.

(Group 1 included 6 laboratories of which 1 used a threshold of  $\geq 5 \text{ mU/L}$  and 5 used a threshold of  $\geq 6 \text{ mU/L}$ ; Group 2 included 3 laboratories, all using a threshold of  $\geq 8 \text{ mU/L}$ ; Group 3 included 4 laboratories, all using a threshold of  $\geq 10 \text{ mU/L}$ ).